昆虫学报 *Acta Entomologica Sinica*, June 2016, 59(6): 622 – 631 doi:10.16380/j.kcxb.2016.06.005

中华按蚊 CPF 家族表皮蛋白基因的全基因组鉴定及其特征分析

刘柏琦, 乔　梁, 许柏英, 郑学令, 陈　斌*

(重庆师范大学生命科学学院, 昆虫与分子生物学研究所, 重庆 401331)

摘要:【目的】鉴定中华按蚊 *Anopheles sinensis* 基因组上的 CPF 家族表皮蛋白基因,分析其基因结构和特征,推测其可能的生物学功能;同时比较研究代表性蚊种的 CPF 家族基因,提供 CPF 家族基因的信息框架。【方法】基于中华按蚊 *An. sinensis*、冈比亚按蚊 *An. gambiae*、微小按蚊 *An. minimus*、埃及伊蚊 *Aedes aegypti*、致倦库蚊 *Culex quinquefasciatus* 和黑腹果蝇 *Drosophila melanogaster* 全基因组序列,以冈比亚按蚊 CPF 家族基因序列为询问序列,采用 BLASTP,TBLASTN 和 HMM 方法鉴定这些物种的 CPF 家族基因;利用生物信息学方法预测中华按蚊 CPF 家族基因的结构、剪切模式、信号肽、跨膜区、结构域和 3D 结构等;采用最大似然法(maximum likelihood, ML)构建这些物种的系统发生关系,推断 CPF 家族基因的起源和进化。【结果】中华按蚊、冈比亚按蚊、微小按蚊、埃及伊蚊、致倦库蚊和黑腹果蝇全基因组共有 4, 4, 4, 3, 3 和 3 个 CPF 家族基因。中华按蚊的 CPF 基因被分别命名为 *AsCPF*1, *AsCPF*2, *AsCPF*3 和 *AsCPF*4,这些 AsCPF 基因的全长 cDNA 序列分别为 736, 2 021, 531 和 1 001 bp,分别编码 219, 345, 148 和 185 个氨基酸。*AsCPF*1, *AsCPF*2 和 *AsCPF*3 仅含有一个内含子,但 *AsCPF*4 含有 3 个内含子,所有内含子均为 0 位内含子。*AsCPF*1, *AsCPF*2, *AsCPF*3 和 *AsCPF*4 分别有 3, 2, 1 和 2 个不同的选择性剪切子。*AsCPF*3 的表达量最高,其次是 *AsCPF*4, *AsCPF*2 和 *AsCPF*1。推测的 AsCPF1, AsCPF2, AsCPF3 和 AsCPF4 的理论分子量分别为 22.86, 36.47, 15.08 和 18.66 kD,等电点分别为 9.08, 8.97, 9.44 和 9.16。AsCPF 家族蛋白含有保守的 44 个氨基酸基序和 C-末端基序;AsCPF1, AsCPF3 和 AsCPF4 具有信号肽,为分泌型蛋白,而 AsCPF2 缺乏信号肽,为非分泌蛋白。二级结构分析显示,4 个 AsCPF 均具有 α-螺旋,无规卷曲和延伸链,只有 AsCPF4 有一段跨膜片段,位于第 5 – 27 位氨基酸。系统发育分析显示,*CPF*3 基因可能是最早分化出来的 CPF 家族基因,*CPF*1 和 *CPF*2 基因可能是同一祖先基因经过一个基因重复事件分化形成的,*CPF*4 基因很可能是按蚊所特有的,是最晚分化出来的 CPF 基因。以冈比亚按蚊为对照,替换率分析显示,中华按蚊 CPF 表皮蛋白的 Ka/Ks 值均小于 1,表现出纯化选择。【结论】对中华按蚊 CPF 家族基因在全基因组上的鉴定和特征分析,及对代表性蚊虫 CPF 家族基因的比较分析,揭示了蚊虫 CPF 家族基因的多样性、结构和氨基酸特征以及起源和进化,这为该家族基因的进一步研究和利用提供了信息基础。

关键词: 中华按蚊; 表皮蛋白; CPF 家族; 保守基序; 进化

中图分类号: Q966　　**文献标识码:** A　　**文章编号:** 0454-6296(2016)06-0622-10

Identification and characterization of the CPF family of cuticular protein genes in the genome of *Anopheles sinensis* (Diptera: Culicidae)

LIU Bai-Qi, QIAO Liang, XU Bo-Ying, ZHENG Xue-Ling, CHEN Bin* (Institute of Entomology and Molecular Biology, College of Life Sciences, Chongqing Normal University, Chongqing 401331, China)

Abstract:【Aim】This study aims to identify the CPF family (CPFs) of cuticular protein genes in *Anopheles sinensis* genome, to analyze their structure and characteristics, to deduce their possible biological functions, and to investigate and compare the CPFs of representative mosquito species so as to

基金项目:"两江学者"计划专项经费; 国家自然科学基金项目(31372265); 国际原子能机构 CRP 项目(18268/R2); 国家科技基础性工作专项重点项目(2015FY210300)

作者简介: 刘柏琦, 男, 1987 年 11 月生, 甘肃定西人, 硕士研究生, 研究方向为昆虫分子生物学, E-mail: 877399740@qq.com

* 通讯作者 Corresponding author, E-mail: bin.chen@cqnu.edu.cn

收稿日期 Received: 2016-03-25; 接受日期 Accepted: 2016-05-18

provide information frame for the family of genes.【Methods】We identified the CPFs in the genomes of *An. sinensis*, *An. gambiae*, *An. minimus*, *Aedes aegypti*, *Culex quinquefasciatus* and *Drosophila melanogaster* using BLASTP, TBLASTN and HMM with *An. gambiae* CPFs as query, predicted the structure and splicing variation of *An. sinensis* CPF gene and the signal peptide, transmembrane region, structural domain and 3D structure of *An. sinensis* CPF proteins using bioinformatics techniques, and constructed phylogenetic relationships using maximum likelihood (ML) method and deduced the origin and evolution of CPFs in these species.【Results】There are 4, 4, 4, 3, 3 and 3 *CPFs* in *An. sinensis*, *An. gambiae*, *An. minimus*, *Ae. aegypti*, *Cx. quinquefasciatus* and *Dr. melanogaster* genomes, respectively. The *CPFs* in *An. sinensis* were named as *AsCPF*1, *AsCPF*2, *AsCPF*3 and *AsCPF*4, respectively. Their full-length cDNA sequences are 736, 2 021, 531, and 1 001 bp, respectively, encoding 219, 345, 148 and 185 amino acids, respectively. *AsCPF*1, *AsCPF*2 and *AsCPF*3 only have one intron, but *AsCPF*4 contains three introns, which all have phase "0". There are 3, 2, 1 and 2 selective spicing variants for *AsCPF*1, *AsCPF*2, *AsCPF*3 and *AsCPF*4, respectively. *AsCPF*3 has the highest expression quantity, followed by *AsCPF*4, *AsCPF*2 and *AsCPF*1. The theoretical molecular weights of AsCPF1, AsCPF2, AsCPF3 and AsCPF4 are 22.86, 36.47, 15.08 and 18.66 kD, and their isoelectric points are 9.08, 8.97, 9.44 and 9.16, respectively. These AsCPFs contain a 44-amino-acid conserved region and C-terminal region, and all are secretory proteins with signal peptide sequences except for AsCPF2 that is non-secretory protein and lacks a signal peptide sequence. All the four AsCPFs have alpha helix, random coil and extended strand, and only AsCPF4 has a transmembrane region that is located between amino acid 5 to 27. Phylogenetic analysis showed that *CPF*3 might be the earliest derived CPF gene, *CPF*1 and *CPF*2 might originate from a common ancestor and consequently experienced a gene duplication event, and *CPF*4 might be unique for *Anopheles* mosquitoes and the latest derived CPF gene. The Ka/Ks ratio of *CPFs* are all less than 1 in *An. sinensis* in reference to *An. gambiae*, suggesting the purification selection of these genes in evolution.【Conclusion】The whole-genome identification and characteristics analysis of CPFs in *An. sinensis* and the comparison of *CPFs* in representative mosquito species revealed the diversity, structure and amino acid characteristics and the origin and evolution of the CPF family of genes in mosquitoes, which provides a comprehensive information frame for further research and utilization of the CPF gene family.

Key words: *Anopheles sinensis*; cuticular protein; CPF family; conservative motif; evolution

昆虫的表皮是昆虫体壁皮细胞分泌物形成的一种高度有序的层状结构，是昆虫适应复杂外界环境的重要保护器官，具有高等动物的皮肤和骨骼的双重功能，在昆虫发育过程中的体型的塑造，水分的维持，抵御外界病原体的攻击和维持正常的活动能力等方面起着重要作用(Delon and Payre, 2004; Moussian *et al*., 2005; Willis *et al*., 2005; 刘喃喃等, 2006; 孙虹霞等, 2007)。昆虫表皮的主要成分是表皮蛋白(cuticular proteins, CPs)和几丁质，表皮蛋白是昆虫重要的结构蛋白。表皮蛋白基因构成一个超基因家族，基于保守的氨基酸基序被进一步分为12个家族，即 CPR, CPF, CPFL, TWDL, CPLCA, CPLCG, CPLCW, CPLCP, CPAP1, CPAP3, CPG 和 Apidermin(He *et al*., 2007; Togawa *et al*., 2007; Willis, 2010)。自 Snyder 等(1982)首次报道了黑腹果蝇 *Drosophila melanogaster* 的4条表皮蛋白基因序列以来，至2014年在 NCBI 中收录的表皮蛋白基因序列已超过1 400条(梁欣等, 2014)。Andersen 等(1997)在黄粉虫 *Tenebrio molitor* 和东亚飞蝗 *Locusta migratoria* 中首次报道了6条 CPF 表皮蛋白序列，其保守基序为一段51个氨基酸的残基，被命名为 CPF 家族；其后，Togawa 等(2007)根据可搜索的表皮蛋白序列，对 CPF 家族的基序进行了修订，发现 CPF 蛋白保守基序只有42～44个氨基酸，且 C-末端保守，其保守基序为：A-(LIV)-x-(SA)-(QS)-x-(SQ)-x-(IV)-(LV)-R-S-x-G-(N/G)-x(3)-V-S-x-Y-(ST)-K-(TA)-(VI)-D-(ST)-(PA)-(YF)-S-S-V-x-K-x-D-x-R-(IV)-(ST)-N-x-(GA)(Togawa *et al*., 2007)。目前，尚未有关于 CPF 家族表皮蛋白基因的全基因组鉴定和生物信息学分析研究，对其功能的研究甚少。Guan 等(2006)在研究果蝇的形体时发现，当表皮蛋白基因 *DmTwlD*1 发生突变后，导致果蝇身体的纵横比例变小，导致"矮胖"，表明 *DmTwlD*1 基因参与昆虫体型的构建；Togawa 等(2007)对冈比亚按蚊 *Anopheles gambiae* 的研究发现，4个 *AgCPF* 基因的 mRNA 只在蛹期或成虫蜕皮前期表达，说明该基因可

能与成虫体壁上表皮的形成相关。

中华按蚊 *Anopheles sinensis* 是我国及东南亚地区疟疾的主要传播媒介，广泛分布于阿富汗、中国、韩国、日本，南至印度尼西亚的广大地区（Sinka *et al*., 2011; Chen *et al*., 2014）。已有研究表明，中华按蚊对多种杀虫剂产生了抗性，主要表现在行为抗性（蚊虫躲避杀虫剂）、表皮抗性（表皮增厚，药物穿透性降低）、代谢抗性（解毒酶活性增强）和靶标抗性（靶标位点的不敏感）4 个方面（Wondji *et al*., 2009; Edi *et al*., 2012）。其中对代谢抗性和靶标抗性已有大量的研究报道，但由于方法学的原因，对表皮抗性和行为抗性一直少有基因水平的研究。近年来，随着生物技术的快速发展，转录组（transcriptome）、基因组（genome）和蛋白质组（proteome）等组学新方法的应用，表皮蛋白基因在表皮的整合、体型的塑造、骨化部位的构建、适应环境的能力及其他生物学等方面的功能研究日益引起人们的关注（Dittmer *et al*., 2012）。如 Reid 等（2012）发现，在致倦库蚊 *Culex quinquefasciatus* 抗菊酯类杀虫剂的品系中有两个上调表达的表皮蛋白基因，分别是 RR2 型和 CPLC 型；在冈比亚按蚊抗拟除虫菊酯杀虫剂品系中，表皮蛋白基因 *CPR*30 表达量显著升高，说明该表皮蛋白基因与杀虫剂抗性有关（Edi *et al*., 2012）。

近年来，重庆师范大学对中华按蚊基因组开展了精细测序和注释工作，并对不同发育时期中华按蚊进行了转录组测序（Chen *et al*., 2014），目前正在开展杀虫剂抗性基因组和功能基因组研究。本研究基于中华按蚊基因组测序数据，采用 BLASTP, TBLASTN 和 Hidden Markov Model（HMM）方法系统地开展了全基因组 CPF 家族基因的鉴定和分类，进而预测了其基因的结构、序列特征、替换率等基因特征，采用同样方法也在冈比亚按蚊、微小按蚊 *An. minimus*、埃及伊蚊 *Aedes aegypti*、致倦库蚊和黑腹果蝇全基因组上鉴定和分类了 CPF 家族的基因，并运用最大似然法（maximum likelihood, ML）构建和讨论了这些昆虫 CPF 家族基因的系统发育和进化，为进一步研究 CPF 家族基因奠定了信息基础。

1 材料与方法

1.1 数据来源

中华按蚊基因组和转录组数据（SRA 登录号：SRA073189）来自于重庆师范大学昆虫与分子生物学研究所，冈比亚按蚊、微小按蚊、埃及伊蚊、致倦库蚊和黑腹果蝇等昆虫的基因组序列下载自 NCBI 的 GenBank 数据库（http://www.ncbi.nlm.nih.gov/）或 VectorBase 数据库（https://www.vectorbase.org/）。

1.2 CPF 家族基因的鉴定和转录

首先，以冈比亚按蚊 CPF 家族氨基酸序列作为查询序列，采用 BLASTP 和 TBLASTN 在中华按蚊基因组数据库中进行同源性搜索，E-value $<1\times10^{-5}$ 作为阈值；其次，采用 HMM（Pfam 号：PF11018）搜索，将得到的候选基因再进行手工校对和系统发育关系分析，选取与冈比亚按蚊 CPF 相似性最高的序列，进一步完成序列的验证。使用鉴定出的 CPF 基因序列作为查询序列，采用 BLASTP 搜索转录组数据库，检测鉴定的 CPF 基因是否转录，选择性的剪切模式。使用标准的 FPKM（fragment per kb per million reads）估计各选择性剪切转录子的表达丰度。

1.3 CPF 家族基因的特征分析

使用 DNAMAN7.0（http://dnaman.software.informer.com/7.0/）鉴定中华按蚊 CPF 家族 cDNA 序列的开放阅读框并翻译成氨基酸序列；使用 BLAST 工具（http://www.ncbi.nlm.nih.gov/BLAST/）进行序列相似性搜索；采用软件 ExPASy（http://www.expasy.org/）预测中华按蚊 CPF 家族蛋白的理论分子量和等电点等；采用 ClustalW 软件（Thompson *et al*., 2002）对中华按蚊 CPF 家族蛋白和其他昆虫同源 CPF 序列进行多重序列比对，并用 Color Align Conservation 软件（http://www.bioinformatics.org/sms2/color_align_cons.html）进行染色；使用 ProtScale 软件（http://www.expasy.org/cgi-bin/protscale.pl）进行蛋白质疏水性分析；使用 TMHMM-2.0（http://www.cbs.dtu.dk/services/TMHMM-2.0/）进行蛋白质跨膜区分析；使用 SignalP 4.1 软件（http://www.cbs.dtu.dk/services/SignalP/）进行信号肽预测；利用蛋白质亚细胞定位软件 TargetP（http://www.cbs.dtu.dk/services/TargetP/）对蛋白进行定位预测；使用软件 NPS（http://npsa-pbil.ibcp.fr）对蛋白质二级结构预测；对于蛋白质三级结构，首先通过 PSI-BLAST 在 PDB（Protein Data Bank）中搜索与中华按蚊 CPF 相似性高的蛋白序列，然后采用 SWISS-MODEL（http://swissmodel.expasy.org/）进行同源建模及 3D 结构预测（唐尧等, 2014）。

1.4 CPF 家族基因系统发育分析

以中华按蚊 CPF 家族序列作为询问序列，采用

同样方法鉴定其他代表性昆虫基因组上的 CPF 基因，基于由基因核酸序列推导的氨基酸序列做系统发育分析，使用黑腹果蝇的 CPF 基因作为外群；利用 MEGA5 软件（Tamura *et al.*, 2011）中的最大似然法构建中华按蚊、冈比亚按蚊、微小按蚊、埃及伊蚊、致倦库蚊和黑腹果蝇 CPF 家族基因序列的系统发育树；使用 1 000 次重复计算系统发育树上的 bootstrap 值（自展分析值），大于 50% 自展值标于系统发育树讨论其系统发育关系。

1.5　CPF 家族基因替换率分析

将中华按蚊 CPF 家族基因的核苷酸序列与冈比亚按蚊同源序列（去除终止密码子）通过 ClustalW 比对，其结果用 KaKs_Calculator 软件（Zhang *et al.*, 2006）计算它们的非同义替换率（Ka），同义替换率（Ks）及 Ka/Ks 比值，讨论 CPF 家族基因的选择压力和选择效应。

2　结果

2.1　中华按蚊 CPF 家族基因的鉴定

通过搜索中华按蚊基因组数据库，得到中华按蚊 4 个候选 CPF 基因，与登录号为 AGAP010900, AGAP010901, AGAP004690 和 AGAP000382 的冈比亚按蚊基因同源性最高，核苷酸序列一致性分别为 88%，83%，77% 和 92%。结构域分析显示这 4 个基因的编码氨基酸均具有昆虫 CPF 家族蛋白保守的基序，即 44 个氨基酸的区域和 C-末端区域（图 1）（Togawa *et al.*, 2007），确信这 4 个基因均为 CPF 家族基因。该 4 条序列的 cDNA 均具有起始密码子（ATG）和终止密码子（TAA），为全长序列，依次命名为 *AsCPF*1，*AsCPF*2，*AsCPF*3 和 *AsCPF*4，其中 *AsCPF*1 推导的氨基酸序列如图 1。中华按蚊 CPF 家

```
-186                                                 TGGCGCAGCGGTAACGCAAGAAAACACTACACCATAGTCGTT
-144 GGATCGAATCCCGAGTCTGGCACTCCCCAGTATTATGAACCACTGACCAACGATCTATCATATAACCGCCTT
-72  TCGGTTACAGAAATTCTCTACGGAGAGCTAGTCCTACTCGGGGATGTTAAGCCAATGAAGAAGAATCATTCG
1    ATGGCATTCAAGTTCGTCGTCTTCCTGGCCTCGTTGGCCGTCGCTAGCGCTGGATACCTGGAGGCTGGCCAT
1     M  A  F  K  F  V  V  F  L  A  S  L  A  V  A  S  A  G  Y  L  E  A  G  H
73   GCCGTCCAGTACGCCGCCCCGGTTGCTCACTACTCGCCGGCTTCGTCGGTGAGCTACAGCACCATCTCGCAG
25    A  V  Q  Y  A  A  P  V  A  H  Y  S  P  A  S  S  V  S  Y  S  T  I  S  Q
145  GCTGCCCCGGCCAAGCTGGCCTACGCCGCCCCGGTTGCCAAGACCATCTCGTACGCCGCCCCTCAGGTGTAC
49    A  A  P  A  K  L  A  Y  A  A  P  V  A  K  T  I  S  Y  A  A  P  Q  V  Y
217  GCCGCCCCGCAGGTCTATGCCGCCCCGGTTGCCAAGACCGTCATCTCCAGCCCGGCCGTCGGTGCCACGCAC
73    A  A  P  Q  V  Y  A  A  P  V  A  K  T  V  I  S  S  P  A  V  G  A  T  H
289  GAGAGCACGATCCGTTCCCATGATGGAACCGTCTCGCACTACTCCAAGGCTGTCGACACCGCTTTCTCGAGC
97    E  S  T  I  R  S  H  D  G  T  V  S  H  Y  S  K  A  V  D  T  A  F  S  S
361  GTCCGCAAGTCGGACACCCGCATCACCAACGAGCTGCCCAAGTACACCTATGCCCAGCCCGTGCTGACCAAG
121   V  R  K  S  D  T  R  I  T  N  E  L  P  K  Y  T  Y  A  Q  P  V  L  T  K
433  CAGGTTGCCTATGCTGCCCCGGCTGTCCACACCACCTATGCTGCTCCGGCCGTCCACACCAGCTATGCTGCC
145   Q  V  A  Y  A  A  P  A  V  H  T  T  Y  A  A  P  A  V  H  T  S  Y  A  A
505  CCGGCTGTCCACACCACCTATGCTGCCCCGGCTGTTGCCACCTACGCCCATGCTGCCCCGGCTGTCCACACC
169   P  A  V  H  T  T  Y  A  A  P  A  V  A  T  Y  A  H  A  A  P  A  V  H  T
577  TCCACCAAGACTCTGACCTACTCGCCGGCCGTCCAGGTTGCGCACACCACCTATGAGGATGCTCATGCCCAT
193   S  T  K  T  L  T  Y  S  P  A  V  Q  V  A  H  T  T  Y  E  D  A  H  A  H
649  TATGCCTGGTAAGCCGGCTGGCTCGTAGCTCAGAGTCTTGGCGGCGACGACGGGAGCGGCATGGGCGTACAC
217   Y  A  W  *
721  CGGGGTGCTTTGGTGGTACAGAGGCTGGTGCTCAGCGACGAGGGTCTTGGTGTAAGTCGGCTGCTGGACGTA
766  CGCGGTCTTGGTGTAGGTCGGCTGCGAGACGTATGCGGTCTTGGT
```

图 1　中华按蚊 *AsCPF*1 cDNA 及其推导的氨基酸序列

Fig. 1　cDNA and deduced amino acid sequence of *AsCPF*1 from *Anopheles sinensis*

图中左边数值为核苷酸和氨基酸序列编号，翻译起始密码子和终止密码子被加框，CPF 保守结构域 44 个氨基酸区域和 C-末端区域用下划线标出。The numbers on the left are the positions of nucleotides and amino acids on the sequences, the start and stop codon are boxed, and the 44-amino-acid conserved region and C-terminal region are underlined.

族基因的基本信息如表 1 所示,包括 scaffold 位置、cDNA 长度、编码区长度、氨基酸大小、AT 和 GC 含量。从表 1 可以看出中华按蚊该家族基因中,*AsCPF*1 和 *AsCPF*2 分布于 scaffold9 上,对应冈比亚按蚊染色体 3L,*AsCPF*3 和 *AsCPF*4 分别分布于 scaffold67 和 scaffold16 上,对应冈比亚按蚊染色体 2L 和 X(Togawa *et al*., 2007),由 148 ~ 345 个氨基酸组成,GC 含量所占比例高,达到 60% 以上。

2.2 中华按蚊 CPF 家族基因的特征

2.2.1 CPF 氨基酸的理化性质:通过 ExPaSy 软件的 ProtParam 分析了该家族成员氨基酸的理论分子量和等电点,结果显示:*AsCPF*1, *AsCPF*2, *AsCPF*3 和 *AsCPF*4 编码蛋白的理论分子量分别为 22.86, 36.47, 15.08 和 18.66 kD,等电点依次为 9.08, 8.97, 9.44 和 9.16。氨基酸组成预测,发现 4 个蛋白中均是丙氨酸(Ala)所占比例最高,依次为 24.2%, 22.9%, 21.6% 和 31.9%,另外脯氨酸(Pro)和缬氨酸(Val)等疏水氨基酸比较丰富,但亲水性的丝氨酸(Ser)和络氨酸(Tyr)含量也较丰富,亲水性氨基酸平均质量分数达 17.33%(表 2)。

表 1 中华按蚊 CPF 表皮蛋白基因的基本信息

Table 1 Basic information of the CPF cuticular protein genes in *Anopheles sinensis*

基因名 Gene name	Scaffold 位置 Scaffold location	cDNA 长度(bp) cDNA length	编码区长度(bp) Coding region length	氨基酸大小 Amino acid size	(A+T)/(G+C) (%)
*AsCPF*1	scaffold9:1118090 - 1118825	736	660	219	35.3/64.7
*AsCPF*2	scaffold9:1112775 - 1114795	2 021	1 038	345	38.2/61.8
*AsCPF*3	scaffold67:1263469 - 1263999	531	447	148	39.4/60.6
*AsCPF*4	scaffold16:11487563 - 11488563	1 001	558	185	32.1/67.8

表 2 中华按蚊 CPF 表皮蛋白的氨基酸组成

Table 2 Amino acid composition of the CPF cuticular proteins in *Anopheles sinensis*

基因名 Gene name	比例 Proportion (%)						
	丙氨酸 Ala	脯氨酸 Pro	丝氨酸 Ser	苏氨酸 Thr	酪氨酸 Tyr	缬氨酸 Val	组氨酸 + 其他 His + others
*AsCPF*1	24.2	7.3	9.6	10.5	8.7	11.0	5.9 + 22.8
*AsCPF*2	22.9	7.8	8.4	9.0	8.1	10.1	5.8 + 27.9
*AsCPF*3	21.6	8.1	8.8	6.1	7.4	9.5	4.1 + 34.4
*AsCPF*4	31.9	8.6	8.6	0.5	9.7	9.7	5.9 + 25.1

2.2.2 CPF 氨基酸的保守结构域:利用 SMART 软件分析中华按蚊 CPF 家族蛋白,对其中的 AsCPF1 氨基酸序列与其他 3 种蚊虫 CPF1 氨基酸序列进行多重序列比对。根据图 2 序列一致性比对结果显示,中华按蚊与其他 3 种蚊虫 CPF1 氨基酸具有相同的保守结构域(图 2)。

2.2.3 CPF 家族基因的结构:基因结构分析表明 *AsCPF*1,*AsCPF*2 和 *AsCPF*3 基因仅含有一个内含子,*AsCPF*4 基因具有 3 个内含子(图 6),所有内含子均为 0 位内含子(内含子位于一密码子的第 3 位核苷酸和另一密码子的第 1 位核苷酸之间)。这与冈比亚按蚊该家族的基因结构基本相同,只是 *CPF*4 基因内含子位相不一致,在冈比亚按蚊中该基因的 3 个内含子有 3 种不同类型的相位,即 0 位内含子、1 位内含子(位于同一密码子第 1 位核苷酸和第 2 位核苷酸之间)和 2 位内含子(位于同一密码子第 2 位核苷酸和第 3 位核苷酸之间)(表 3)。

表 3 中华按蚊和冈比亚按蚊 CPF 家族基因内含子-外显子结构

Table 3 Intron and exon organization of the CPF family genes in *Anopheles sinensis* and *An. gambiae*

基因名 Gene name	外显子长度 (bp) Exon size	内含子长度 (bp) Intron size	内含子相位 (0/1/2) Intron phase
*AsCPF*1	12/648	76	0
*AsCPF*2	204/834	983	0
*AsCPF*3	9/438	84	0
*AsCPF*4	12/261/195/84	90/281/78	0/0/0
*AgCPF*1	12/705	79	0
*AgCPF*2	48/876	224	0
*AgCPF*3	91/557	67	0
*AgCPF*4	12/265/175/85	303/96/70	0/1/2

2.2.4 信号肽预测和亚细胞定位:SignalP4.1 软件预测结果显示,除 AsCPF2 的蛋白序列不存在信号肽外,其他 3 个蛋白序列都具有信号肽;AsCPF1 和

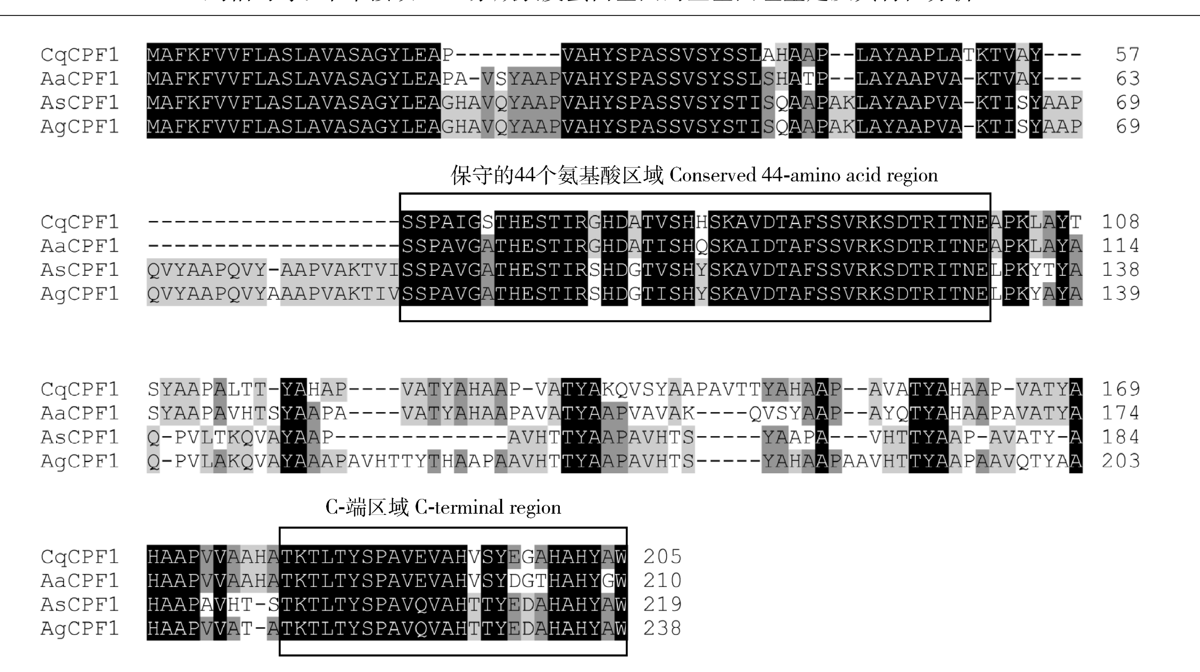


图 2　中华按蚊 *CPF*1 基因与其他按蚊 *CPF*1 基因氨基酸序列比较

Fig. 2　Amino acid sequence comparison of *AsCPF*1 and *CPF*1 genes from other *Anopheles* species

Cq：致倦库蚊 *Culex quinquefasciatus*；Aa：埃及伊蚊 *Aedes aegypti*；As：中华按蚊 *Anopheles sinensis*；Ag：冈比亚按蚊 *Anopheles gambiae*. 图中用线框标出的为 CPF 家族的 2 个保守结构域；黑色、灰色和白色阴影分别表示氨基酸序列保守性为 100%，80% 和 80% 以下。Two conserved domains are line-boxed. Black, grey and white shade denote the amino acids with 100%, 80% and below 80% identity, respectively.

AsCPF4 表皮蛋白的信号肽序列均位于第 1－17 位氨基酸，而 AsCPF3 表皮蛋白在第 1－16 位氨基酸。TargetP1.1 分析显示，AsCPF1，AsCPF3 和 AsCPF4 蛋白均为分泌蛋白，即分泌到细胞周质，故该蛋白定位到胞外；而 AsCPF2 蛋白无信号肽，为非分泌蛋白。

2.2.5　疏水性预测：通过 ExPaSy 软件的 ProtScale 功能预测 CPF 家族蛋白的疏水性，结果显示 CPF 家族蛋白均为亲水性蛋白，无疏水区域存在。

2.2.6　跨膜区分析：采用 TMHMM Server v.2.0 软件预测该家族蛋白的跨膜区，发现只有 AsCPF4 有一个跨膜片段，位于第 5－27 位氨基酸之间，且该蛋白可能为膜结合蛋白（图 3）。

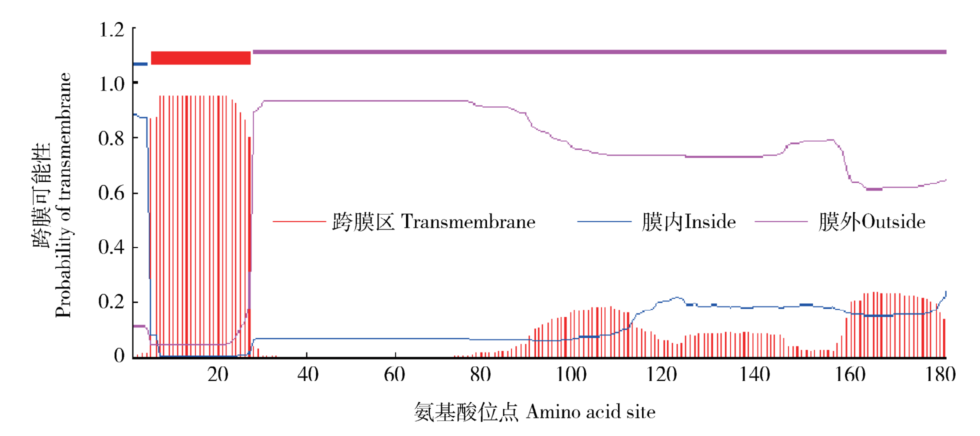


图 3　中华按蚊 CPF4 蛋白跨膜区拓扑模型

Fig. 3　Deduced transmembrane domain topology model of AsCPF4 protein in *Anopheles sinensis*

2.2.7　二级结构和三级结构预测：NPS 软件预测结果显示，AsCPF1 蛋白二级结构中 α-螺旋，无规卷曲和延伸链分别占 36.06%，45.66% 和 18.26%，在 AsCPF2 蛋白中则分别为 37.39%，46.67% 和 15.94%，在 AsCPF3 蛋白中分别占 41.22%，35.81% 和 22.97%，在 AsCPF4 蛋白中分别占 67.57%，26.49% 和 5.95%。可以看出在 AsCPF1 和 AsCPF2 表皮蛋白中无规卷曲占比最高，而在 AsCPF3 和 AsCPF4 中 α-螺旋占比最高。由此推测，无规卷曲是 AsCPF1 和 AsCPF2 二级结构中最大量的结构元件，α-螺旋和延伸链分散于整个蛋白质中；而在 AsCPF3 和 AsCPF4 的二级结构中，α-螺旋是最大量的结构元

件。通过 PSI-BLAST 搜索,只有 AsCPF2 和 AsCPF3 分别与 NMDA 受体蛋白(PDB 编号: 4tlm. 1. A)和孢子蛋白(PDB 编号: 4u5a. 4. A)氨基酸序列一致性最高,分别为 15.63% 和 33.33%,被选作同源建模的模板,再通过 SWISS-MODEL 建模预测得到 AsCPF2 和 AsCPF3 蛋白的三级结构(图 4)。

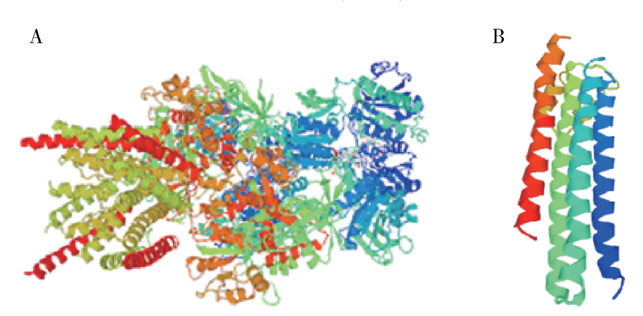


图 4 中华按蚊 AsCPF2(A)和 AsCPF3(B)的三维结构图
Fig. 4 Predicted 3D structure of AsCPF2 (A) and AsCPF3 (B) in *Anopheles sinensis*

2.3 CPF 家族基因系统发育

经过全基因鉴定,中华按蚊、冈比亚按蚊、微小按蚊、埃及伊蚊、致倦库蚊和黑腹果蝇分别有 4, 4, 4, 3, 3 和 3 个 CPF 家族基因。所研究的 5 种蚊虫的 *CPF*3 基因形成了一个独立的支系,具有 100% 的 bootstrap 值支持,该支系位于系统发育树的底端,与黑腹果蝇 3 个 CPF 基因形成的支系形成姊妹群,黑腹果蝇也具有 *CPF*3 基因,这表明 *CPF*3 基因可能是最早分化出来的 CPF 基因(图 5)。所研究蚊虫的 *CPF*1 和 *CPF*2 基因聚集在同一分支上,而且同一蚊种的 *CPF*1 和 *CPF*2 基因部分形成独立的分支,黑腹果蝇也具有 *CPF*1 和 *CPF*2 基因,表明 *CPF*1 和 *CPF*2 近缘,可能来自于同一祖先基因,随后经历了基因的重复事件。中华按蚊、冈比亚按蚊和微小按蚊的 *CPF*4 聚集在一个独立支系,具有 100% 的 bootstrap 支持,所研究的其他 3 个种缺乏 *CPF*4 基因,很可能 *CPF*4 基因是按蚊属蚊虫特有的,是最晚分化出来的 CPF 基因。

2.4 CPF 家族基因的替换率分析

非同义替换率(Ka)和同义替换率(Ks)的比值

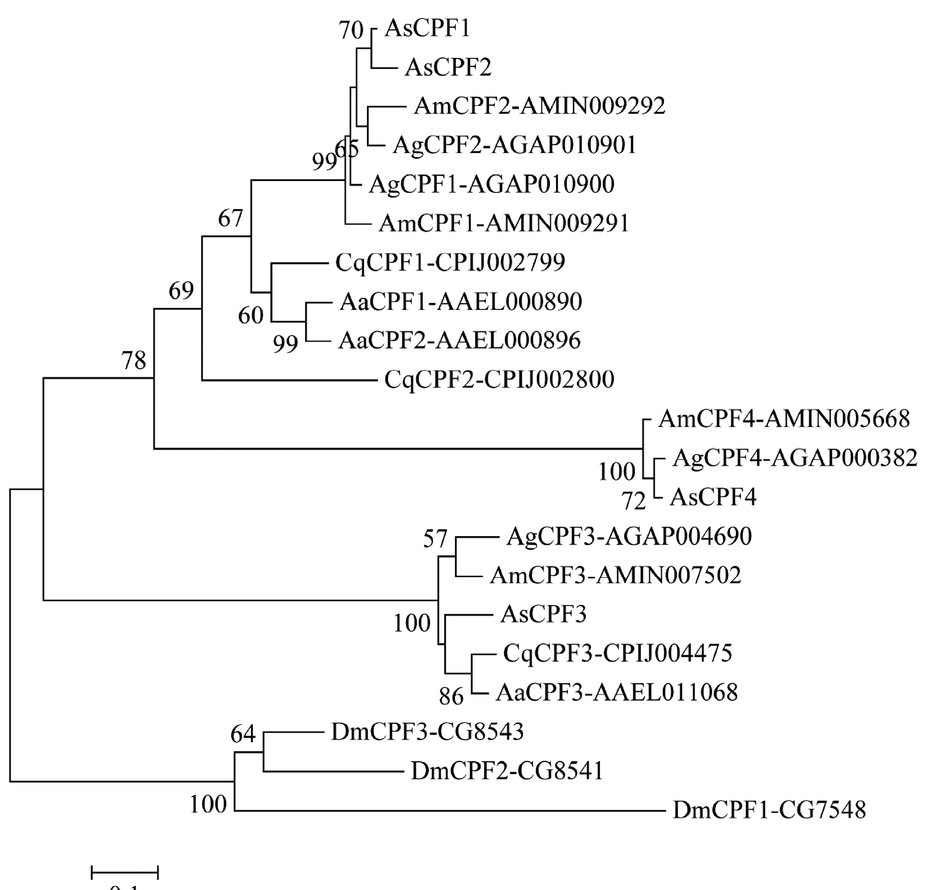


图 5 基于推导的氨基酸序列的 6 种昆虫 CPF 家族基因的系统发育关系(最大似然法)
Fig. 5 The phylogenetic relationships of the CPF family genes of 6 insect species based on the deduced amino acid sequences (Maximum likelihood)
As: 中华按蚊 *Anopheles sinensis*; Cq: 致倦库蚊 *Culex quinquefasciatus*; Aa: 埃及伊蚊 *Aedes aegypti*; Ag: 冈比亚按蚊 *Anopheles gambiae*; Am: 微小按蚊 *Anopheles minimus*; Dm: 黑腹果蝇 *Drrosophila melanogaster*. 各序列缩写的种名后为序列的 GenBank 登录号;大于 50% 的 bootstrap 值标记在树的分支节点上;标尺代表系统发育距离。Following the sequence abbreviation names are their GenBank accession numbers, percentage bootstrap values higher than 50% are marked on each branch, and the scale bar indicates the phylogenetic distance.

可以判断蛋白编码基因是否存在选择压力，也可以反映该基因的保守程度。以冈比亚按蚊为对照，中华按蚊 4 个 CPF 表皮蛋白基因的 Ka，Ks 及 Ka/Ks 比值如表 4 所示，Ka/Ks 值均小于 1，介于 0.02～0.13之间，表现出纯化选择，说明该家族基因的进化压力大，相对十分保守。其中，*AsCPF2* 基因的 Ka/Ks 值高于其他基因，表明 *AsCPF2* 经受的选择压力较小。

表 4　中华按蚊 CPF 表皮蛋白基因的 Ka，Ks 及 Ka/Ks

Table 4　The Ka, Ks and Ka/Ks values of the CPF cuticular protein genes in *Anopheles sinensis*

基因名 Gene name	Ka	Ks	Ka/Ks
CPF1	0.0284568	1.14286	0.0248996
CPF2	0.0839516	0.673782	0.124598
CPF3	0.0880845	2.12132	0.0415234
CPF4	0.016147	0.451267	0.0357815

Ka/Ks＞1，正选择；Ka/Ks＝1，中性选择；Ka/Ks＜1，纯化选择。Ka/Ks＞1, positive selection; Ka/Ks＝1, neutral selection; Ka/Ks＜1, purifying selection.

2.5　CPF 家族基因的剪切性转录和表达

使用各 CPF 基因序列搜索转录组数据库，结果表明所有 4 个 AsCPF 基因都有转录子，*AsCPF1*，*AsCPF2*，*AsCPF3* 和 *AsCPF4* 分别有 3，2，1 和 2 个不同的选择性剪切子（表 5）。图 6 显示 CPF 家族各基因的结构及不同选择性剪切的位置。

以 FPKM 作为选择性剪切子的表达丰度标准，*AsCPF1* 的 FPKM 为 5.7735（CL1630.Contig5_5）～40.3858（CL1630.Contig3_5），*AsCPF2* 为 43.1798（CL1630.Contig2_5）～53.5612（CL1630.Contig1_5），*AsCPF3* 为 384.9348（Unigene14355_5），*AsCPF4* 为 0.3832（CL1336.Contig2_5）～116.2549（CL1336.Contig1_5）（表 5）。计算各基因的所有剪切子 FPKM 总和可以看出，*AsCPF3* 的表达量最大（FPKM＝384.9348），其次是 *AsCPF4*（FPKM＝116.6381），*AsCPF2*（FPKM＝96.741）和 *AsCPF1*（FPKM＝84.4908）（表 5）。

表 5　中华按蚊 CPF 表皮蛋白基因的选择性剪切及表达丰度

Table 5　Splicing variants and transcription richness of the CPF cuticular protein genes in *Anopheles sinensis*

基因名 Gene name	选择性剪切 ID Splicing variants ID	表达丰度 FPKM	表达丰度总和 Total FPKM
AsCPF1	CL1630.Contig3_5	40.3858	84.4908
	CL1630.Contig4_5	38.3315	
	CL1630.Contig5_5	5.7735	
AsCPF2	CL1630.Contig1_5	53.5612	96.741
	CL1630.Contig2_5	43.1798	
AsCPF3	Unigene14355_5	384.9348	384.9348
AsCPF4	CL1336.Contig1_5	116.2549	116.6381
	CL1336.Contig2_5	0.3832	

FPKM: Number of fragments per kilobase of exon model per million mapped reads.

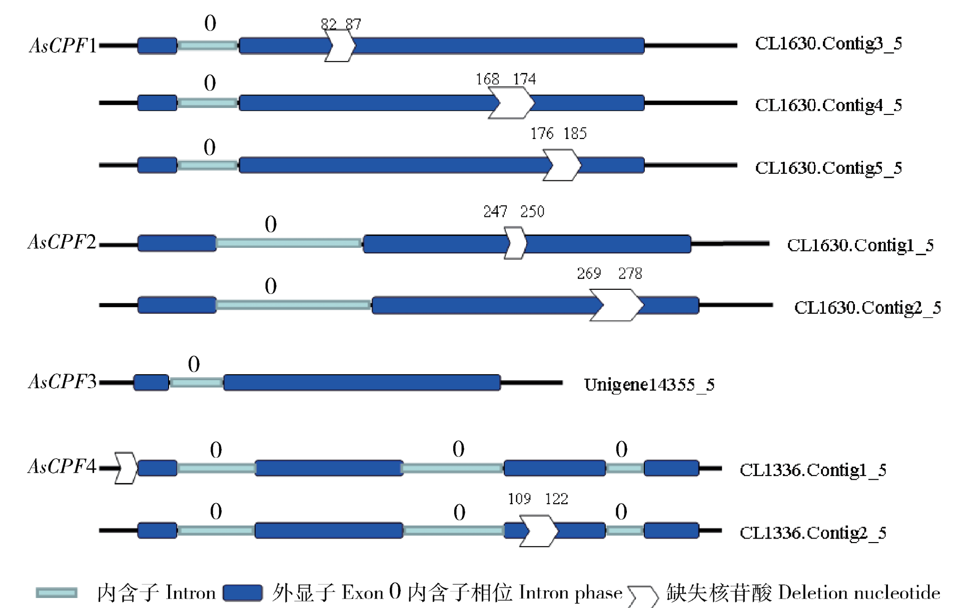


图 6　中华按蚊 CPF 表皮蛋白基因的结构及选择性剪切

Fig. 6　Structure and splicing variants of the CPF cuticular protein genes in *Anopheles sinensis*

3　讨论

表皮蛋白是组成昆虫表皮外骨骼的主要成分之一，表皮蛋白基因的数量约占基因组中蛋白编码基因总数的 2%（Willis, 2010; Neafsey *et al*., 2015）。目前已在多种昆虫中鉴定出 CPF 家族基因，如在冈比亚按蚊中该家族有 4 个基因成员（GenBank 登录

号：EF382662），黑腹果蝇中有 3 个成员（GenBank 登录号：NM_139874，NM_139870 和 NM_139869）（Togawa *et al.*，2007）。本研究基于中华按蚊基因组数据，共鉴定得到 4 个中华按蚊 CPF 家族基因，与已报道的冈比亚按蚊 CPF 家族基因数量相同。利用生物信息学方法全面分析了这些基因的序列特征，发现 *AsCPF*1，*AsCPF*2，*AsCPF*3 和 *AsCPF*4 分别有 3，2，1 和 2 个不同的选择性剪切子，*AsCPF*3 的表达量最大，其次是 *AsCPF*4，*AsCPF*2 和 *AsCPF*1。同时在冈比亚按蚊、微小按蚊、埃及伊蚊、致倦库蚊和黑腹果蝇全基因组上分别鉴定出了 4，4，3，3 和 3 个 CPF 家族的基因，并对中华按蚊、冈比亚按蚊、微小按蚊、埃及伊蚊、致倦库蚊和黑腹果蝇 CPF 家族基因进行了比较分析，丰富了昆虫表皮蛋白超家族基因数据，有利于推动表皮蛋白在蚊虫生长发育中基因表达、调控、功能及其他生物学方面的研究，如表皮蛋白基因在昆虫表皮的发生、分化、整合、体形的塑造，个体行为及活动能力以及先天免疫等生理现象和生理过程中的重要作用（梁欣等，2014），同时，也为基于表皮蛋白基因的害虫防控策略研究提供基础信息。

与其他昆虫 CPF 基因编码的氨基酸序列进行同源性比对发现，中华按蚊 CPF 家族基因均具有该家族典型的 2 个保守结构域，即 44 个氨基酸局域和 C-末端局域，这种典型的保守结构域暗示这些基因可能与其特定的生物学功能有关（Andersen *et al.*，1995）。中华按蚊 4 个 CPF 基因的基因结构与冈比亚按蚊基本一致，只有 *AsCPF*4 和 *AgCPF*4 基因的内含子相位不一致（Holt *et al.*，2002），这种不变的基因结构是否与功能相关，还有待进一步的研究。跨膜区预测显示，只有 AsCPF4 为膜结合蛋白，由于膜结合蛋白通常不溶于水，分离纯化比较困难，且不易成晶体，很难确定其结构。系统发育关系显示 *CPF*3 基因可能是最早分化出来的 CPF 基因，*CPF*1 和 *CPF*2 基因间的序列相似性最高，可能是同一祖先基因经过一个基因重复事件分化形成的，*CPF*4 基因很可能是按蚊属蚊虫特有的，是最晚分化出来的 CPF 基因。

在自然界中，非同义替换一般都是有害突变，在这些突变位点上，碱基的替换将由于负选择作用而保持比较低的突变速率（周琦和王文，2004）。为了确定 CPF 表皮蛋白基因在进化上的选择模式，利用 Ka 与 Ks 的比值来评估，如 Ka/Ks 值 <1，则认为有纯化选择的压力，即同义替换的速率高于非同义替换的速率，Ka/Ks 值越小，表明该基因承受的选择压力越大，保守程度越高（Wagner，2002；周琦和王文，2004）。中华按蚊 CPF 表皮蛋白基因的 Ka/Ks 值均远小于 1，介于 0.02 ~ 0.13 之间，表明该家族表皮蛋白基因均为纯化选择，进化上相对保守，暗示这些表皮蛋白对蚊虫的生存和特定的功能是必不可少的（梁九波等，2008）。

冈比亚按蚊中研究发现，CPF 表皮蛋白基因仅仅在蛹或成虫蜕皮前表达，参与上表皮的形成（Togawa *et al.*，2007；Papandreou *et al.*，2010），我们推测中华按蚊中该家族基因的功能可能与冈比亚按蚊相似，但具体的生物学功能及其他领域的应用还有待我们进一步的探究。

参考文献（References）

Andersen SO, Hojrup P, Roepstorff P, 1995. Insect cuticular proteins. *Insect Biochem. Mol. Biol.*, 25: 153 – 176.

Andersen SO, Rafn K, Roepstorff P, 1997. Sequence studies of proteins from larval and pupal cuticle of the yellow meal worm, *Tenebrio molitor*. *Insect Biochem. Mol. Biol.*, 27(2): 121 – 131.

Chen B, Zhang YJ, He Z, Li W, Si F, Tang Y, He Q, Qiao L, Yan Z, Fu W, Che Y, 2014. *De novo* transcriptome sequencing and sequence analysis of the malaria vector *Anopheles sinensis* (Diptera: Culicidae). *Parasit. Vectors*, 7: 314.

Delon I, Payre F, 2004. Evolution of larval morphology in flies: get in shape with shavenbaby. *Trends Genet.*, 20(7): 305 – 313.

Dittmer NT, Hiromasa Y, Tomich JM, Lu N, Beeman RW, Kramer KJ, Kanost MR, 2012. Proteomic and transcriptomic analyses of rigid and membranous cuticles and epidermis from the elytra and hindwings of the red flour beetle, *Tribolium castaneum*. *J. Prot. Res.*, 11(1): 269 – 278.

Edi CV, Koudou BG, Jones CM, Weetman D, Ranson H, 2012. Multiple-insecticide resistance in *Anopheles gambiae* mosquitoes, Southern Côte d'Ivoire. *Emerg. Infect. Dis.*, 18(9): 1508 – 1511.

Guan X, Middlebrooks BW, Alexander S, Wasserman SA, 2006. Mutation of TweedleD, a member of an unconventional cuticle protein family, alters body shape in *Drosophila*. *Proc. Natl. Acad. Sci. USA*, 103(45): 16794 – 16799.

He N, Botelho JM, Mcnall RJ, Belozerov V, Dunn WA, Mize T, Orlando R, Willis JH, 2007. Proteomic analysis of cast cuticles from *Anopheles gambiae* by tandem mass spectrometry. *Insect Biochem. Mol. Biol.*, 37: 135 – 146.

Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R, Salzberg SL, Loftus B, Yandell M, Majoros WH, Rusch DB, Lai Z, Kraft CL, Abril JF, Anthouard V, Arensburger P, Atkinson PW, Baden H, de Berardinis V, Baldwin D, Benes V, Biedler J, 2002. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science*, 298(5591): 129 – 149.

Liang JB, Liu BL, Zhan ZG, He NJ, 2008. Bioinformation analysis of cuticular protein genes in the silkworm, *Bombyx mori*. *Science of Sericulture*, 34(3): 405 – 416. ［梁九波, 刘碧朗, 占智高, 何宁佳, 2008. 家蚕表皮蛋白基因的生物信息学分析. 蚕业科学, 34(3): 405 – 416］

Liang X, Chen B, Qiao L, 2014. Research progress in insect cuticular protein genes. *Acta Entomologica Sinica*, 57(9): 1084 – 1093. ［梁欣, 陈斌, 乔梁, 2014. 昆虫表皮蛋白基因研究进展. 昆虫学报, 57(9): 1084 – 1093］

Liu NN, Zhu F, Xu Q, Pridgeon JW, Gao XW, 2006. Behavioral change, physiological modification, and metabolic detoxification: mechanisms of insecticide resistance. *Acta Entomologica Sinica*, 49(4): 671 – 679. ［刘喃喃, 朱芳, 徐强, Pridgeon JW, 高希武, 2006. 昆虫抗药性机理: 行为和生理改变及解毒代谢增强. 昆虫学报, 49(4): 671 – 679］

Moussian B, Schwarz H, Bartoszewski S, Nusslein-Volhard C, 2005. Involvement of chitin in exoskeleton morphogenesis in *Drosophila melanogaster*. *J. Morphol*., 264(1): 117 – 130.

Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, Amon J, Arcà B, Arensburger P, Artemov G, Assour LA, Basseri H, Berlin A, Birren BW, Blandin SA, Brockman AI, Burkot TR, Burt A, Chan CS, Chauve C, Chiu JC, Christensen M, Costantini C, Davidson VL, Deligianni E, Dottorini T, Dritsou V, Gabriel SB, Guelbeogo WM, Hall AB, Han MV, Hlaing T, Hughes DS, 2015. Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. *Science*, 347(6217): 1258522.

Papandreou NC, Iconomidou VA, Willis JH, Hamodrakas SJ, 2010. A possible structural model of members of the CPF family of cuticular proteins implicating binding to components other than chitin. *J. Insect Physiol*., 56(10): 1420 – 1426.

Reid WR, Zhang L, Liu F, Liu NN, 2012. The transcriptome profile of the mosquito *Culex quinquefasciatus* following permethrin selection. *PLoS ONE*, 7(10): e47163.

Sinka ME, Bangs MJ, Manguin S, Chareonviriyaphap T, Patil AP, Temperley WH, Gething PW, Elyazar IR, Kabaria CW, Harbach RE, Hay SI, 2011. The dominant *Anopheles* vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic précis. *Parasit. Vectors*, 4: 89.

Snyder MP, Kimbrell D, Hunkapiller M, Hill R, Fristrom J, Davidson N, 1982. A transposable element that splits the promoter region inactivates a *Drosophila* cuticle protein gene. *Proc. Natl. Acad. Sci. USA*, 79(23): 7430 – 7434.

Sun HX, Liu Y, Zhang GR, 2007. Effects of heavy metal pollution on insects. *Acta Entomologica Sinica*, 50(2): 178 – 185. ［孙虹霞, 刘颖, 张古忍, 2007. 重金属污染对昆虫生长发育的影响. 昆虫学报, 50(2): 178 – 185］

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S, 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol*., 28(10): 2731 – 2739.

Tang Y, Qiao L, Zhang YJ, Che YF, Hong R, Chen B, 2014. Identification and bioinformatics analysis of genes of the CYP6Y subfamily in *Anopheles sinensis* (Diptera: Culicidae). *Acta Entomologica Sinica*, 57(6): 663 – 672. ［唐尧, 乔梁, 张玉娟, 车艳飞, 洪瑞, 陈斌, 2014. 中华按蚊 CYP6Y 亚家族基因的鉴定和生物信息学分析. 昆虫学报, 57(6): 663 – 672］

Thompson JD, Gibson TJ, Higgins DG, 2002. Multiple sequence alignment using ClustalW and ClustalX. In: Current Protocols in Bioinformatics, Chapter 2, Unit 2.3.

Togawa T, Augustine Dunn WA, Emmons AC, Willis JH, 2007. CPF and CPFL, two related gene families encoding cuticular proteins of *Anopheles gambiae* and other insects. *Insect Biochem. Mol. Biol*., 37(7): 675 – 688.

Wagner A, 2002. Selection and gene duplication: a view from the genome. *Genome Biol*., 3(5): reviews 1012.

Willis JH, 2010. Structural cuticular proteins from arthropods: annotation, nomenclature, and sequence characteristics in the genomics era. *Insect Biochem. Mol. Biol*., 40(3): 189 – 204.

Willis JH, Iconomidou VA, Smith RF, Hamodrakas SJ, 2005. Cuticular proteins. In: Gilbert LI, Latrou K, Gill SS eds. Comprehensive Molecular Insect Science. Vol. 4. Elsevier, Oxford. 79 – 109.

Wondji CS, Irving H, Morgan J, Lobo NF, Collins FH, Hunt RH, Coetzee M, Hemingway J, Ranson H, 2009. Two duplicated P450 genes are associated with pyrethroid resistance in *Anopheles funestus*, a major malaria vector. *Genome Res*., 19(3): 452 – 459.

Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, 2006. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics*, 4: 259 – 263.

Zhou Q, Wang W, 2004. Detecting natural selection at the DNA level. *Zoological Research*, 25(1): 73 – 80. ［周琦, 王文, 2004. DNA 水平自然选择作用的检测. 动物学研究, 25(1): 73 – 80］

（责任编辑：袁德成）